

Deep Learning Course - Analisis Regresi Dataset (MLP Model)

Muhammad Naufal Firdaus_1103220083

1. Jika menggunakan model MLP dengan 3 hidden layer (256-128-64) menghasilkan *underfitting* pada dataset ini, modifikasi apa yang akan dilakukan pada arsitektur? Jelaskan alasan setiap perubahan dengan mempertimbangkan *bias-variance tradeoff*!

Jika MLP dengan 3 layer tersebut menunjukkan *underfitting* Ketika training, maka kemungkinan model tersebut terlalu sederhana sehingga membuat bias terlalu tinggi dan tidak mampu untuk menangkap kompleksitas pola dalam data. Maka dapat dilakukan :

a. Menambah layer atau neuron

- Dengan menambahkan layer atau neuron, maka arsitektur menjadi lebih kompleks dan bisa menangkap pola non-linear yang lebih rumit
 - Menambah kapasitas model mengurangi bias tapi meningkatkan variance
- Mengurangi regularisasi (Jika menggunakan)
 - Regularisasi yang terlalu kuat bisa memaksa model terlalu sederhana dan memperparah *underfitting*
- Ganti aktivasi yang tadinya menggunakan ReLU menjadi LeakyReLU/ELU
 - Aktivasi ini merupakan alternatif ,engatasi dying ReLU dan mempercepat konvergensi

2. Selain MSE, loss function apa yang mungkin cocok untuk dataset ini? Bandingkan kelebihan dan kekurangannya, serta situasi spesifik di mana alternatif tersebut lebih unggul daripada MSE!

Beberapa alternatif untuk Loss Function (dalam kasus ini regresi) selain MSE bisa menggunakan MAE, Huber Loss, dan Log-cosh Loss. Perbandingan serta situasi tersebut lebih unggul adalah demikian :

Loss Function	Kelebihan	Kekurangan	Cocok Saat
MSE	Peka terhadap outlier, penalti besar untuk kesalahan besar	Terlalu peka terhadap outlier	Target bersih dan terdistribusi normal
MAE	Robust terhadap outlier	Tidak terdiferensiasi di nol (sulit untuk optimisasi)	Banyak outlier di target
Huber	Gabungan MAE dan MSE: sensitif untuk kesalahan kecil, robust untuk outlier	Memiliki parameter δ yang harus disetel	Ketidakseimbangan error skala sedang
Log-Cosh	Mirip MSE tapi tidak penalti besar, smooth	Sedikit lebih mahal secara komputasi	Distribusi target non-normal dan banyak noise

3. **Jika salah satu fitur memiliki range nilai 0-1, sedangkan fitur lain 100-1000, bagaimana ini memengaruhi pelatihan MLP? Jelaskan mekanisme matematis (e.g., gradien, weight update) yang terdampak!**

Jika ada fitur dengan Range 0-1 sedangkan fitur lain 100-1000, maka ada ketidakseimbangan dan menyebabkan beberapa efek matematika, diantaranya adalah :

- a. Update weight (via backpropagation) menjadi tidak seimbang:
 - Gradien untuk fitur dengan nilai besar akan mendominasi penurunan loss.
 - Fitur dengan range kecil bisa diabaikan sehingga menyebabkan slow learning atau no learning.
- b. Vanishing / exploding gradients:
 - Karena aktivasi (misal sigmoid) bisa jenuh akibat input terlalu besar.

Untuk mengatasi hal tersebut, bisa menggunakan feature scaling (misal StandardScaler, MinMaxScaler, atau RobustScaler) agar semua fitur berada dalam rentang yang sama. dengan begitu akan menjaga kestabilan gradient descent dan meningkatkan efisiensi optimisasi.

4. **Tanpa mengetahui nama fitur, bagaimana Anda mengukur kontribusi relatif setiap fitur terhadap prediksi model? Jelaskan metode teknikal (e.g., permutation importance, weight analysis) dan keterbatasannya!**

Ketika tidak mengetahui nama fitur (pada kasus ini saya mengubah baris pertama menjadi target, feat_1, feat_2, dst), kita tetap bisa mengukur kontribusi relatif masing-masing fitur terhadap prediksi model menggunakan beberapa metode teknis :

a. **Weight Analysis (khusus MLP)**

- **Konsep:**
 - Dalam MLP, tiap input terhubung ke layer pertama dengan bobot w_i .
 - Besar bobot (absolute) bisa menunjukkan kontribusi fitur.
 - Dapat dijumlahkan atau dirata-rata antar neuron.
- **Kelebihan:**
 - Cepat, karena tidak perlu re-training atau evaluasi ulang.
 - Mudah diakses dari `.weights` atau `.get_weights()`.
- **Keterbatasan:**
 - Tidak mempertimbangkan **aktivasi non-linear** dan interaksi fitur.
 - Bisa misleading kalau data belum dinormalisasi.
 - **Feature scaling memengaruhi** interpretasi.

b. **SHAP (SHapley Additive exPlanations)**

- **Konsep:**
 - Berdasarkan teori permainan: berapa kontribusi “pemain” (fitur) terhadap “skor akhir” (prediksi)?
 - Cocok untuk regresi maupun klasifikasi.
 - Menjelaskan prediksi individual dan global.

- **Kelebihan:**
 - Sangat interpretable dan akurat.
 - Cocok untuk model non-linear seperti MLP, XGBoost.
 - Bisa menjelaskan interaksi antar fitur.

- **Keterbatasan:**
 - Komputasi mahal untuk dataset besar.
 - Implementasi agak kompleks (pakai shap library).
 - Bisa sulit dibaca tanpa visualisasi.

5. Bagaimana Anda mendesain eksperimen untuk memilih learning rate dan batch size secara optimal? Sertakan analisis tradeoff antara komputasi dan stabilitas pelatihan!

Pada dataset ini, untuk melakukan eksperimen kearning rate dan batch size yang optimal, digunakan Hyperparameter Tuning untuk mencari desain optimal.

- Tuning digunakan :
 - Grid Search / Random Search dengan kombinasi:
 - Learning rate: [1e-4, 1e-3, 1e-2]
 - Batch size: [32, 64, 128, 256]
 - Bayesian Optimization (dengan Keras Tuner) untuk efisiensi.

Pertimbangan Tradeoff :

Parameter	Nilai Kecil	Nilai Besar
Learning Rate	Stabil tapi lambat konvergen	Cepat, tapi rawan overshoot / divergen
Batch Size	Noisy update sehingga regularisasi alami	Smooth update tapi rawan overfitting